# Description

## FIELD OF THE INVENTION

The present invention relates to a method and system for monitoring public health status with information technology, and more particularly to the early detection of unusual public health events through the analysis of the over-the-counter (OTC) pharmaceutical sales data. The present invention could be directly applied to the implementation of public health decision support systems in the area of bio-intelligence. Another direct application of this invention could be pharmaceutical supplies planning and inventory. A potentially useful application is in providing the workload adjustment for the public health systems and pharmaceutical industries.

## BACKGROUND OF THE INVENTION

Recently, a few pharmaceutical sales surveillance systems have been developed for monitoring public health status. Those systems count the number of sales of categorized medical items, and plot those values over time. Public health experts need to review the trend of categorized medicine sales, map it into a relationship with the public health status, and do data interpretation. Those systems have no functionalities to directly detect unusual public health events before the clinical diagnosis is performed or to directly explain the relationship to public health status through OTC medicine sales data. Because of the above reasons, the detection of unusual public health events and the identification of public health status usually is delayed until a time when the number of patients seeking professional medical help reaches an abnormal level, and the number of confirmed disease cases is above a pre-defined threshold value. This means many people have already been infected, sometimes even the secondary spread of a communicable disease would be underway.

As there is a rise in threats from emerging infectious diseases and a degradation in the quality of the environment, there is an urgent need for a method and system with automated processes to detect unusual public health events faster and more efficiently than through the clinics. This early detection could greatly aid public health workers and even save peoples' lives. Methods and systems with the capacity to systematically identify the public health status from the varying OTC medicine sales data and other early indicators could greatly benefit public health for disease prevention and control. It will also help pharmacy stores in planning and inventory control incorporating the seasonal adjustment.

The system and methods presented herein integrate database technologies, knowledge based techniques, statistical analysis methods, dynamic systems theory and rule systems. The system is an integrated decision support system designed for both the public health and pharmaceutical industry.

Herein, the applied database technology is used for the replicated sales data repository, utilized in organizing the processed data, and for querying and retrieving information. Additionally, the knowledgebase technique approach is utilized in deriving knowledge from data processing, then storing and organizing the knowledge in multiple dimensions along space and time, followed by inference utilizing the rule systems.

4

The state-space form, as a part of the mathematical system theory (Kalman et al 1969), is adapted here in modeling categorized public health status. A category of the public health status, for example, can be the gastrointestinal disease syndrome, the respiratory disease syndrome, or the children flu like diseases. A set of state variables are defined here to represent the categorized public health status, and the state transition mechanisms are developed for modeling the change of public health status over time. The major difference in this described modeling of state transitions presented from the conventional ones is that the transitions here are governed by rule systems. The rule systems evaluate the states and the inputs then make the decision; while in the other systems the state transitions are determined by an explicit algebra function, for example, linear algebra in most cases.

A rule system is a set of rules, arguments, constraints, relations, and responses. A rule can be numerical, logical or both. A hybrid rule system consists of both explicit functions and logical rules. The presented system is a hybrid rule system.

There are no publications found for an OTC pharmaceutical sales surveillance system that include knowledge acquisition on public health. Additionally, up to now, no publication is found for a state-space model with state transitions determined by rule systems in the public health area.


## SUMMARY OF THE INVENTION

The invention consists of the mathematical model describing the change of categorized health status, and more importantly it has the detection methods for unusual health status before evidence appears in clinics, and a dynamic model for the change of a categorized public health status from OTC medicine sales at a specific location. First, the measurement scheme is defined. Then, at a specified place, reference lines are established from historical data. Those reference lines represent the normal values and extreme values of the OTC daily sales during a particular time interval at a specific location. Current daily values, for the same category, are measured by their deviations from the reference lines. Measurements include the relative change, the n-days-cumulated change and the rate of the changes.


### *The setting of the measurement scheme*

### The measurement of the OTC medicine daily sales at a place and a time
A study place can be a store service area, a zip-code area, a city, a county or it can be statewide. The approach is the same for all the areas. Here, to simplify the description, we just state it as 'a study area' or 'at each geographical level'.

A time unit can be a day, a week, a month, a season, or simply x-number of days. The approach is the same for all time units. Here, to simplify the description, we just state it as 'a time unit' .


### Calculations of the reference lines
The main purpose of this method is to detect the irregular change of public health status from the regular change of OTC sales. First, the m-years-historical data, from the current date back to at least one previous year, are processed to derive the reference lines. Next, a time unit is defined for the specified time interval, the averaged time-unit-value is calculated for each category as equation (1) at each geographical level. For example, it could be a monthly-averaged-daily-value for the medicines for gastrointestinal symptoms for each city. Similarly, the standard deviation of the daily

sales is calculated by equation (2), and the confidence interval upper limit of the mean is calculated by equation (3). The results of those three equations could yield the center and the upper reference lines at each corresponding geographical level. Since it is computed in each time interval, the seasonal variations of disease syndromes are maintained; and the computations at each location in different geographical levels portrait the spatial characteristics. For a specified place, the calculations are performed as equation (1), (2) and (3).

Equation (1) yields the center reference line or baseline, while equation (2) and equation (3) yield the upper reference lines, for example, they could be 2-sigma and 3-sigma lines when $t_{mxn-1} = 2$ and $3$. To illustrate the usage of equation (1) to (3), a sample set of The OTC data are plotted in Figure 1. The sample data are the daily OTC sales related to gastrointestinal diseases, three-month long, both in last year and in this year at a study area. For the non-disclosure of business data, the sales amounts are not displayed in the graph. Figure 2 shows the derived monthly reference lines from the sample data in last year, based on equation (1) to (3).

To measure the change of OTC sales at a specified place, or possibly the abnormality, the following equations define three structural components derived from the time-unit sales data. In the following example, the time-unit is daily. The first structural component is defined by equation (4). In a specified place, Equation (4) is the measurement of the relative deviation for daily sales from the center reference line, which is the averaged-daily-value in that time interval and it stands for a normal status unless there was a record of a large-scale outbreak that had happened in the past used to calculate this reference line. Using the calculation in equation (4), if $m$ is small (less than three), there is no requirement for the population data in the specified area unless there is a significant change of the population in one or two years at that place. This is another advantage of this approach.

The second component is defined by equation (5), it is the $n$-days-cumulated-change of the categorized OTC daily deviation from the baseline. The example shows the 7-days when $n$ is from $0$ to $6$. By defining $n$ is greater than 6, Equation (5) smoothes the sales variation in weekdays and on weekends. The physical meaning in Equation (5) is that most of the time the purchased medicine is used in several days, thus, its effects remain for several days. In the application, the value of $n$ is also determined by the categorized medicines. Equation (6) reflects the daily change of the relative deviation, it is a leading indicator of the trend, and this is the third structural component. In the equation, L denotes the current year.

The calculated results from Equation (1), (2), (3), (4), (5) and (6) quantitatively describe the historical daily sales in a normal situation, and the differences of current daily sales from it and the change of those differences. Those calculated results are the base of the supporting sets for the input rule system. Figure 3 illustrates the transformed results by using the same sample data set plotted in Figure 1, based on the equation (1), (4), (5) and (6).

The confidence supporting sets of above components, ($d$, $w$, $v$,), are defined from the cumulated distribution functions as shown in equation (7), (8) and (9).

For example, if $\alpha=0.05$ is specified for month $j$ at a study place, then in the historical data set where years are ($l<L$), the cumulated distribution function $F(d_{l<L,j,i})$ is structured first; next, the supp(($\alpha(k)$) is found as the set of $d_{l<L,j,i}$ such that its cumulated distribution function $F(d_{l<L,j,i}) > (1-0.05)$. Similarly equation (8) and (9) define the confidence supporting sets of the other two components, and Figure 4 to Figure 6 illustrate them graphically.

6

## A Dynamic Model of The Categorized Public Health Status

The dynamic change of the public health status with space and time is modeled here in a new state-space form. This newly invented state-space form differs from the other conventional state-space approach in that here the state transition, input mapping and output mapping are governed by the rule systems; while the conventional state-space form uses crisp algebra or linear algebra in most cases. With a state space notation, at a specified place, the categorized public health status is explicitly modeled by a set of state variables, which are varying over time. Defined by this model, at a specific time, a categorized health status is one of the following: healthy status ($S_h$), critical status ($S_c$), starting-unusual status ($S_s$), upward-trend-unusual status ($S_u$), peak-unusual status ($S_p$), downward-trend-unusual status ($S_d$), and ending-unusual status ($S_e$). The state transitions over the time reflect the dynamic change of the public health status.

The state space **S** is defined with its state variables $\{S: S_h, S_c, S_s, S_u, S_p, S_d, S_e\}$. A validated state transition from state $S_i(k)$ to state $S_j(k+1)$ is determined by the rule systems which operate in relational algebra on its supporting set $X_i(k)$. The validated state transitions are defined in Figure 7 with the arrow arcs, or by Table 1. In Table 1 a zero stands for an invalidated transition, while the validated transition from state $S_i(k)$ to state $S_j(k+1)$ is determined by a rule base $R_{i,j}$, which evaluates the inputs $X_i(k)$ at state $S_i(k)$.

Since in most cases we work on the daily base in the current year, to simplify the notation, the subscript '$L$' (which stands for current year) is omitted in the following equations.

Equation (10) defines the general form of state transitions from a state $S_i(k)$ to another state $S_j(k+1)$, and the transition is determined by a rule system $R_{i,j}$ operating on the supporting set of $X_i(k)$. The quantitative measurement of the state $S_i(k)$ is also defined by Equation (10).

As time advances, for example a time unit can be daily, the state transition from state $S_i(k)$ to state $S_j(k+1)$ is determined by the rule system $R_{i,j}$ which evaluates the supporting set $X_i(k)$, as shown in Equation (10), where $\otimes$ stands for the inference operation, or a rule system operation, which can be logical operations or algebra operations or mixture of them. Equation (10) also defines the value of a state $S_i(k)$ is proportional to the n-days-cumulated deviation in that category at the specified place. The $k_w$ can be defined from the supporting set data, for example it can be related to the threshold value obtained from the historical data set.

Equation (11) describes that, at a state $S_i(k)$, there is the supporting set $X_i(k)$ with 3 structural components, and their thresholds $(\alpha(k), \beta(k), \delta(k))$ can be incorporated and the rule system $B_{i,m}$ maps the components into supporting set $X(k)$. Where $\otimes$ stands for the inference operation, or a rule system operation.

Equation (12) describes the output mapping, which interprets the outputs from a set of states or a state history with the specified weight for the states by $\{\gamma_0(k), \gamma_1(k), ..., \gamma_n(k)\}$. In addition, the rule system combines the background information, $G_i$, such as the environment factors, the population or the age grouped population in the study area.

Equation (13) and (14) define the supporting system $X$ is an additive combination of supporting sets. It means the inputs can be multiple data sources.

Equation (11), (12) and (13) together define the thresholds of the supporting sets. The threshold values are derived from the historical data in the same time interval (e.g. the *jth* month in past *m* years), for the time unit *i* (e.g. the *ith* day), for each component. In equation (11), (12) and (13), the supporting sets are derived from the cumulated probability distributions, $F_d(\alpha)$, $F_w(\beta)$ and $F_v(\delta)$, of the three components.

Equation (15) defines that the value of an output is the combination of the likelihood index of abnormality $(Q_{i,h})$, the trend indicator $(T_{i,h})$ and the potential impact index $(P_{i,h})$. As an example they have been defined here as

$\{Q_{i,h} : \text{(low, medium, high)} \}$,
$\{T_{i,h} : \text{(stable, upward, downward)} \}$,
$\{P_{i,h} : \text{(minor, moderate, significant)} \}$.

For example, a $Y_i = (Q_{i,2}, T_{i,2}, P_{i,3})$ stand for the medium likelihood abnormality, with upward trend status and possible significant potential impact. In reality, this situation might require specified extensive management. In the case where the population density is used to describe the potential impact, $P_i$ is defined as (light density, low density, high density).

The following example is the best mode presently contemplated for carrying out the invention. This description is not to be taken in a limiting sense, but is made merely for the purpose of describing the principles of the invention. The scope of the invention should be determined with reference to the claims. In this example, the OTC sales data has one-year historical records. To make it easier to illustrate, the study geographical area is the store service area which is estimated by the driving distance; and the time unit is daily, the time interval for the computation of reference lines is 30 days, the plotted time period is 91 days (date converted to Day 1 to Day 91), which actually were data from three different months. For the purpose of non-disclosing the real sales data, the sales counts are not displayed in the following examples.

The distribution of structural component 1 for the same month last year can be used to derive the threshold value for $d_{i,1}(k)$. For example, if an $\alpha$-level is set as $\alpha(k) = 0.05$, then the $\alpha$-support of $X_{i,1}(k) = 75\%$ by equation (11). Similarly, Figure 5 and Figure 6 are illustrations of deriving threshold values for Component 2 and Component 3 by equation (12) and equation (13).

## Summary

The invented dynamic system model and the data analysis methods have been described. The measurement scheme is defined first. The reference lines are derived from the historical data at the same geographical unit. Next, the incoming data are transformed into three structured components as defined by equation (4), (5) and (6). The dynamic system model for public health status was developed in the form of state space as equation (10), (11) and (12). There are seven state variables are defined, with the validated state transitions in Table 1. The state transitions

8

are determined by the rule systems. The system supporting sets are compiled from the transformed incoming data, while the system outputs are mapped from the state variables' history and possible other data sources.

The set of figures have illustrated the principles and computational methods. The brief descriptions of the graphics are summarized as followings.

Figure 1 shows the examples of the categorized OTC daily sales in three months from both last year and this year in the same study area;

Figure 2 shows the derived reference lines from the historical data set with equation (1), (2) and (3), here they are from the last year's data;

Figure 3 is a graph of three structural components computed by equation (4), (5) and (6);

Figure 4 illustrates how to derive the confidence supporting set for Component 1 by equation (7);

Figure 5 illustrates how to derive the confidence supporting set for Component 2 by equation (8);

Figure 6 illustrates how to derive the confidence supporting set for Component 3 by equation (9);

Figure 7 is the diagram of defined state variables and the validated state transitions;

Figure 8 is the example to illustrate the rule system by using the transformed results of incoming OTC daily sales.

Figure 9 is the example of the OTC sales abnormality analysis in supporting the risk assessment and management. The map displays the derived service areas, the area population density, and the categorized OTC sales analysis.

Table 1 shows the validated state transitions.

Equation (1) defines the calculation of the central reference line at a specified place.

Equation (2) defines the calculation of the deviation from the central reference line.

Equation (3) defines the calculation of the upper reference line.

Equation (4) defines the calculation of the relative deviation of the incoming daily data from the central reference line.

Equation (5) defines the calculation of the n-days-cumulated-deviation of the incoming data from central reference line.

Equation (6) defines the calculation of the change of the relative deviation.

Equation (7) defines the confidence supporting set of the first component.

Equation (8) defines the confidence supporting set of the second component.

Equation (9) defines the confidence supporting set of the third component.

Equation (10) defines the system state transitions and the measurement of the state.

Equation (11) defines the system input mapping from the supporting sets and their threshold values are incorporated.

Equation (12) defines the system outputs are mapped from the state history and the background information can be incorporated.

Equation (13) defines the supporting space.

Equation (14) defines the supporting system is an additive combination of supporting sets.

Equation (15) defines the values of an output are the combination of a likelihood index, a trend indicator and a potential impact index.

## An example of the change of categorized health status as the state transitions with the rule system operations

To further illustrate the method and system disclosed here, the examples of the state transitions, and rule systems operations are provided. Those examples are not the complete rule operations, but merely used for the purpose of illustration.

The example data set is the daily OTC medicine sales of gastrointestinal (GI) diseases in a studied area, with one year's historical data. The base line is computed by equation (1), and transformed three structural components by equation (4), (5) and (6). Figure 8 displays the transformed daily OTC medicine sales of gastrointestinal diseases in the studied area. The data from Day 33 to Day 77 are taken to illustrate the defined state transitions for the estimation of public health status in the category of gastrointestinal diseases in the studied place.

In this example, the *a-level* is set as 0.05, thus, the threshold values for three components are 75%, 150% and 70% (corresponding to Figures 4, 5 and 6). Here the coefficient $k_w$ in equation (10) is defined in reference of the threshold value. Thus,

$K_w = 1 / 150.$

Based on equation (10) to equation (15), the system state transitions from Day 33 to Day 77 can be summarized as the followings.

Day 33 to Day 50, the status in state $S_h(k; k= 33, .., 50)$ (healthy status);
Day 51 the state transited to state $S_c(51)$ (critical status);
Day 52 the state transited to $S_s(52)$ (starting unusual status);
Day 53 the state transited to $S_u(53)$ (upward trend unusual status), and it is above the threshold value of *w(k)*;
Day 54 to Day 55, the states are $S_u(k ; k = 54, 55)$ (upward trend unusual status);
Day 56 the states in $S_u(56)$ (upward trend unusual status) and the component 1 is above the threshold value of *a(k)*;
Day 57 to Day 59, the states are $S_u(k ; k = 57, ..., 59)$ (upward trend unusual status);
Day 60 the state is $S_p(60)$ (peak in the unusual);
Day 61 to Day 64 the states in $S_d(k; k=61, 62, ..., 64))$ (downward trend unusual);
Day 65 the states in $S_d(65)$ (downward trend unusual) while component 1 is above the threshold value of *a(k)*;
Day 66 to Day 71 the states in $S_d(k; k=66, 67, ..., 71))$ (downward trend unusual);
Day 72 the states in $S_d(72)$ and it is below the threshold value of *w(k)*;
Day 73 to Day 75 the states in $S_d(k; k=73, 74, 75))$ (downward trend unusual);
Day 76 the state is $S_e(76)$ (ending unusual);
Day 77 the state back to $S_h(77)$ (healthy status).

To further illustrate the method and system disclosed here, some examples of the state transitions with the rule system operations are provided. Those examples are not the complete rule operations, but merely used for the purpose of illustration. In the following example, the output's potential impact index is assumed 'moderate', $P_{i,2}$.

**Examples of state transition and inputs / outputs with data referenced in Figure (5) to (8):**

Day 33 to Day 50, the status in state $S_h(k; k= 33, .., 49)$ (the healthy status):

$R_{i,j}$ :  If $S(k-1) = S_h$
  and
    $\{ w_i(k-1) < 0 \}$
  then
    $S(k) => S_h$ and its value is $S_h(k) = k_w\, w_i(k) < 0.$

$H_{i,n}$:  If $\max\{S(k), S(k-1), .., S(k-n)\} < 0,$
  Then
      $Y(k) = (Q_{i,1}, T_{i,1}, P_{i,2}).$

Day 51 the state transited to state $S_c(51)$ (a critical status):

$R_{i,j}$ :  if $S(k-1=50) = S_h$
  And
    $\{d_i(k) \in \text{supp}(\alpha(k)) \text{ and } v_i(k) \in \text{supp}(\delta(k))\}$
  then
    $S(k=51) => S_c$, and its value is $S_c(k) = k_w\, w_i(k) = 0.5.$

$H_{i,n}$:  If $\{S(k-1), .., S(k-n)\} = S_h$
  And $S(k) = S_c$
    Then
      $Y(k) = ( Q_{i,2}, T_{i,2}, P_{i,2}).$

Day 52 the state transited to $S_s(52)$ (starting unusual status):

$R_{i,j}$ :  if $S(k-1=51) = S_c$
  And
    $\{d_i(k) \in \text{supp}(\alpha(k)) \text{ and } w_i(k) \in \text{supp}(\beta(k))\}$
  then
    $S(k=52) => S_s$, and its value is $S_s(k) = k_w\, w_i(k) = 0.65.$

$H_{i,n}$:  If $\{S(k-1)\} = S_c$
  And $S(k) = S_s$
    Then
      $Y(k) =( Q_{i,2}, T_{i,2}, P_{i,2}).$

Day 52 to Day 59, the states are $S_u(k ; k = 52, ..., 59)$ (upward-trend in unusual status):

$R_{i,j}$ :  if $S(k-1=52) = S_s$
  And  {
    $\{ d_i(k) \in \text{supp}(\alpha(k)) \text{ and } w_i(k) \in \text{supp}(\beta(k)) \}$
    or
    $\{ d_i(k) \in \text{supp}(\alpha(k)) \text{ and } v_3(k) > 0 \}$
    }
  then
    $S(k) => S_u$, and its value is $S_u(k) = k_w\, w_i(k)$ *(greater than 1.0).*

$H_{i,n}$:  If $\{S(k-1)\} = ( S_c \text{ or } S_u )$

And $S(k) = S_u$
    Then
        $Y(k) = (Q_{i,3}, T_{i,2}, P_{i,2})$.

(Note: at k= 53, $w_i(k)$ reaches the threshold value, this issues an alert in the application.)

......

Day 61 to Day 75 the states in $S_d(k; k=61, 62, ..., 75))$ (downward trend unusual):
$R_{i,j}$ :    if $S(k-1) = S_p$ or $S_d$
        And  {
        { $d_i(k) < 0$  or $v_i(k) < 0$ }
            or
                { $d_i(k) \notin supp(\alpha_1(k))$ or $v_i(k) \notin supp(\beta_1(k))$ }
        }
    then
        $S(k) => S_d$, and its value is $S_d(k) = k_w \, w_i(k)$.
$H_{i,n}$:    If {$S(k-1)$} = ( $S_p$ or $S_d$ )
        And $S(k) = S_d$
        Then
                $Y(k) = = (Q_{i,2}, T_{i,3}, P_{i,2})$.

Day 76 the state is $S_e(76)$ (ending unusual):
$R_{i,j}$ :    if $S(k-1) = S_d$        And
        { $w_i(k) < 0$  or $w_i(k) = 0$ }
    then
        $S(k) => S_e$, and its value is $S_e(k) = k_w \, w_i(k)$.

$H_{i,n}$:    If {$S(k-1)$} = ( $S_d$ )
        And $S(k) = S_e$
        Then
                $Y(k) = = (Q_{i,1}, T_{i,1}, P_{i,2})$.

Day 77 the state transitioned back to $S_h(77)$ (healthy status):
$R_{i,j}$ :    if $S(k-1) = S_e$ or $S_h$        And
        { $w_i(k) < 0$  or $w_i(k) = 0$ }
    then
        $S(k) => S_h$, and its value is $S_h(k) = k_w \, w_i(k)$.

$H_{i,n}$:    If {$S(k-1)$} = ( $S_d$ or $S_e$ )
        And $S(k) = S_h$
        Then
                $Y(k) = (Q_{i,1}, T_{i,1}, P_{i,2})$.

**Example of the system outputs with the GIS map:**

13

The system outputs as described above contain the information of the likelihood of abnormality and the potential impact indicators. The estimated likelihood abnormality is mapped to the health status in the studied area. As an example, the service areas are derived with the driving distance within 5 minutes and 10 minutes to the stores, and the populations within the service areas are the part of the referenced potential impacts. By using geographical information systems (GIS) with the application of the disclosed methods, Figure 9 is the example that displays the spatial distributions of the likelihood of abnormality and potential impacts. In Figure 9, the abnormality is classified as 'low, stable' for outputs are the likelihood is low and the trend is stable; similarly the 'medium, upward' is for likelihood is medium and trend is upward, and 'high, upward' for likelihood is high and trend is upward. Through the overlay of the OTC sales abnormality with the population density in the service area, a combined output is obtained, Figure 9 is a map display of equation (12).

While the invention herein disclosed has been described by means of specific applications thereof, numerous modifications and variations could be made thereto by those skilled in the art without departing from the scope of the invention set forth in the claims.
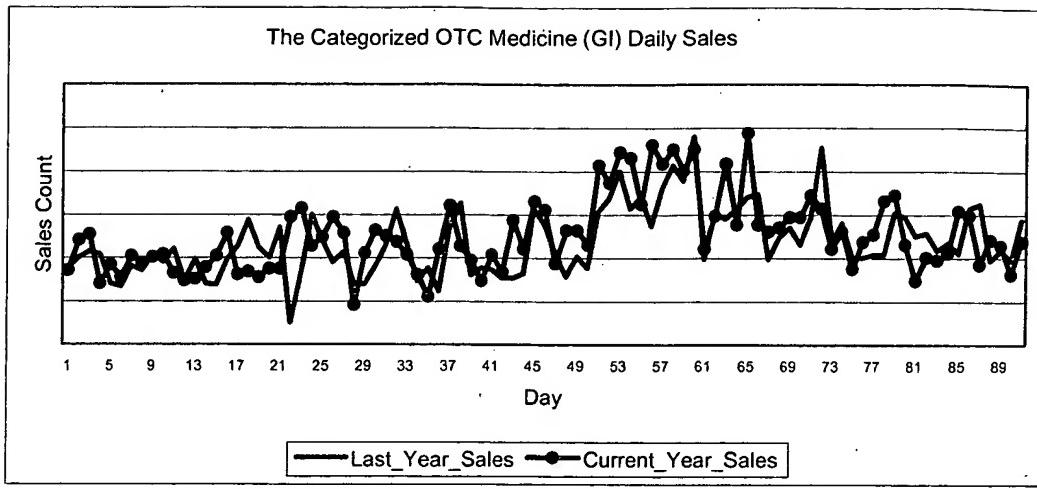
Figure 1.The OTC daily sales in this year and last year in the study area

The Categozied OTC Medicine (GI) Daily Sales
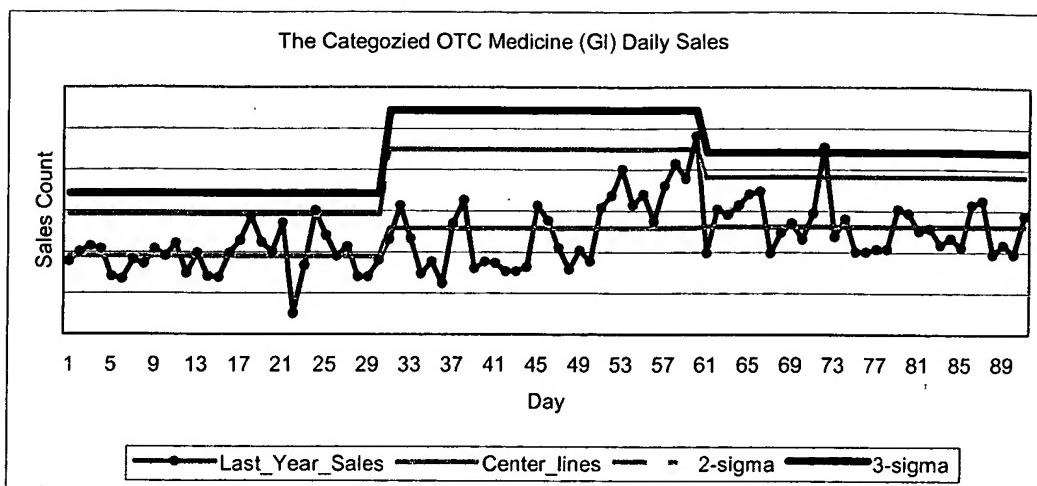
Figure 2.The derived reference lines from the last year's daily sales

The Transformed Results In 3 Components Of OTC Medicine (GI) Daily Sales
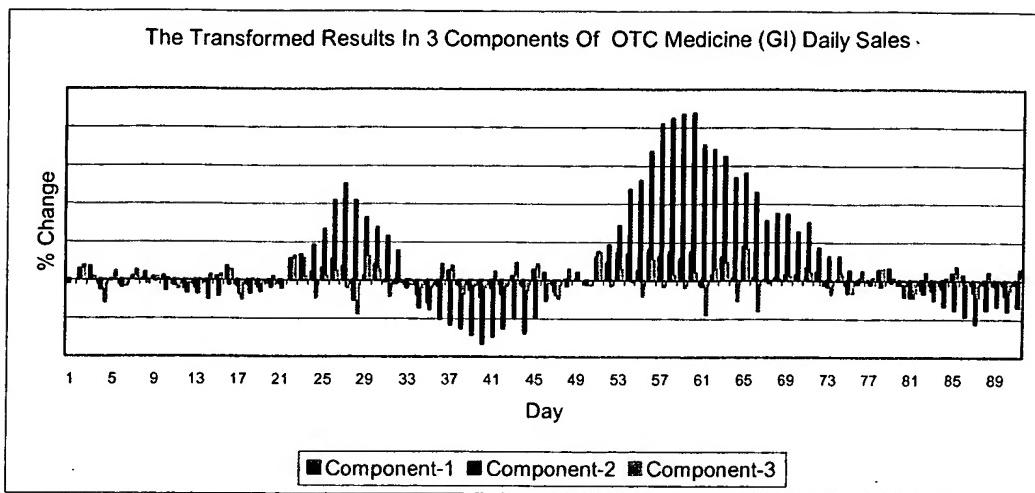
Figure 3.The transformed incoming daily sales in 3 components by Eq. (4), (5) and (6)
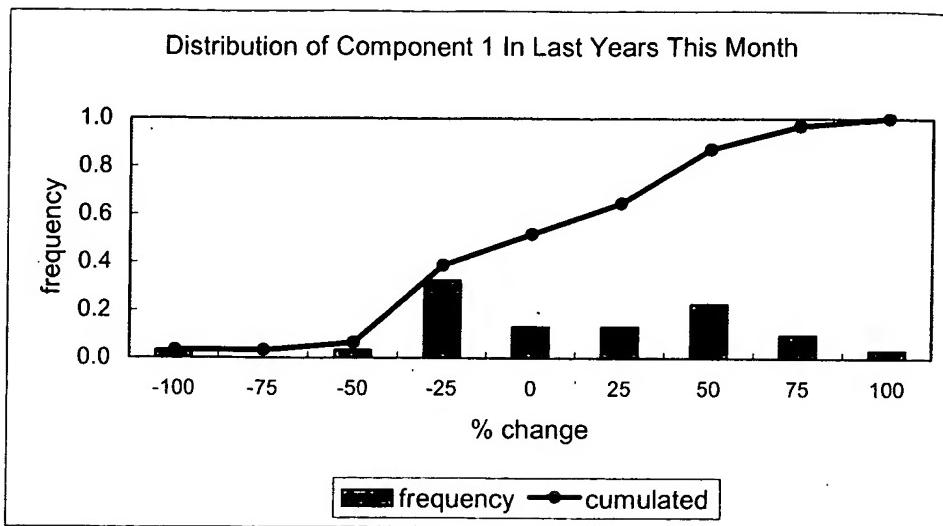
Figure 4.Deriving threshold value from the distribution of Component 1

**Distribution of Component 2 In Last Years This Month**
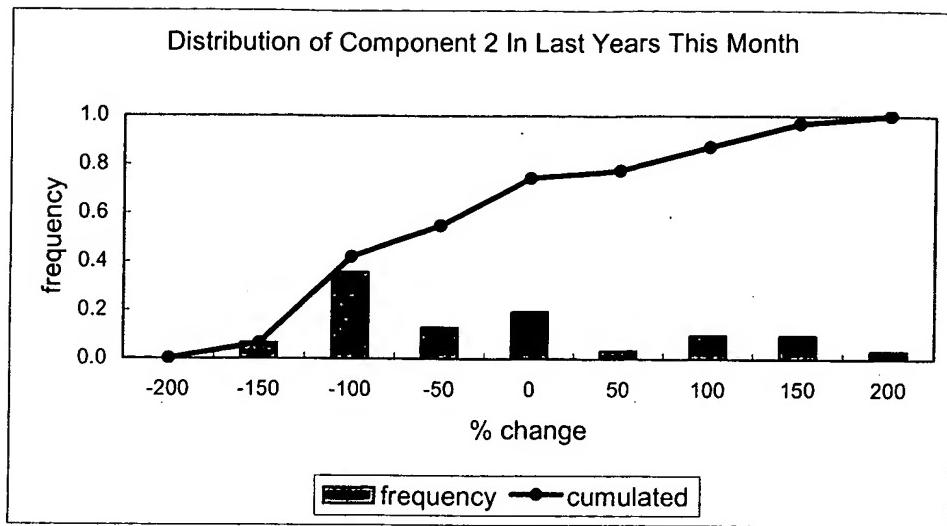
**Figure 5.    Deriving threshold value from the distribution of the component 2**

**Distribution of Component 3 In Last Years This Month**
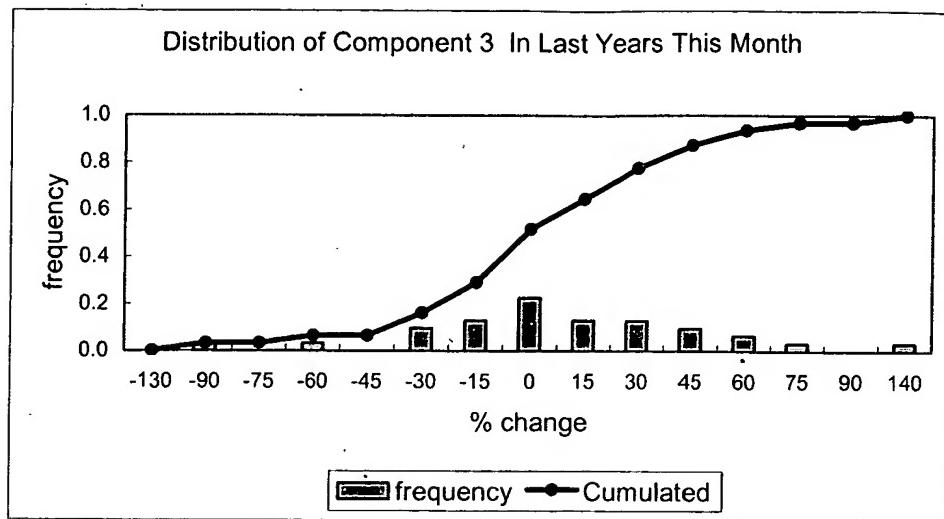
**Figure 6.    Deriving threshold value from the distribution of the component 3**
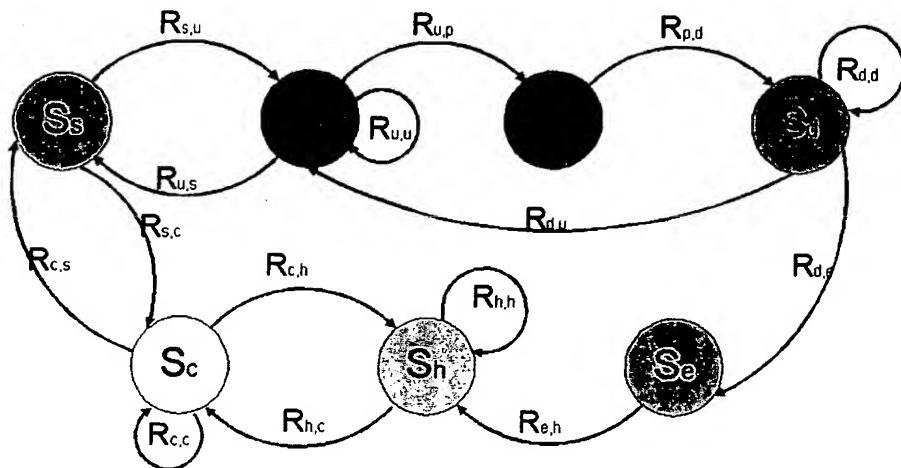
Figure 7.The States and State Transition Diagram of Public Health Status

The Transformed Results In 3 Components Of OTC Medicine (GI) Daily Sales
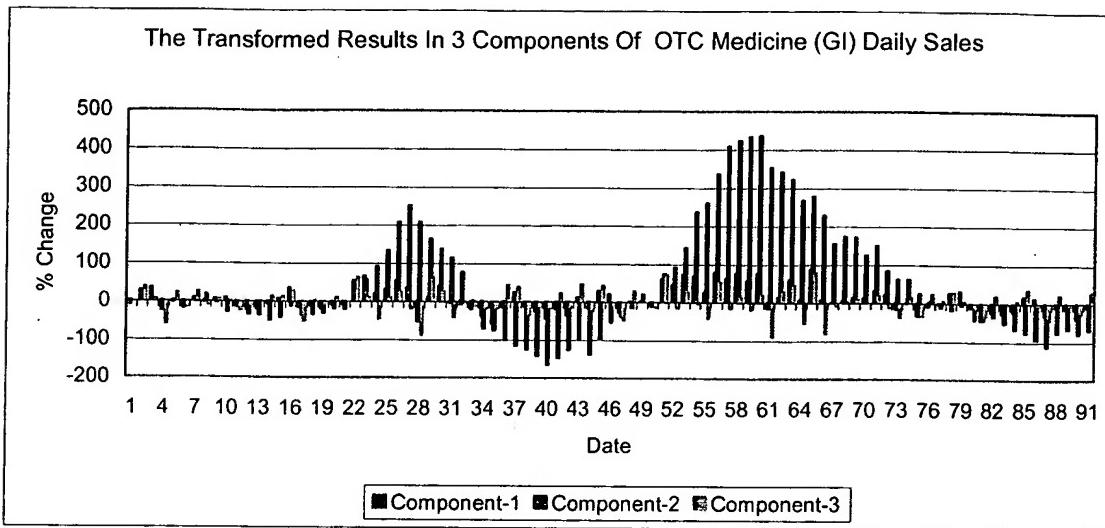
Figure 8.The example of transformed results of OTC daily sales data

The Selected Pharmacy Stores
With The OTC Sales Abnormality
And Populations In Service Areas

Sales Abnormality & Trend
• low, stable
• medium, upward
• high, upward
Roads
Population Density In Zone 1
1000 - 5000
5000 - 10000
10000 - 15000
15000 - 20000
20000 - 40000
Populations Density In Zone 2
5000 - 20000
20000 - 40000
40000 - 60000
60000 - 80000
80000 - 200000
Zip Code Areas
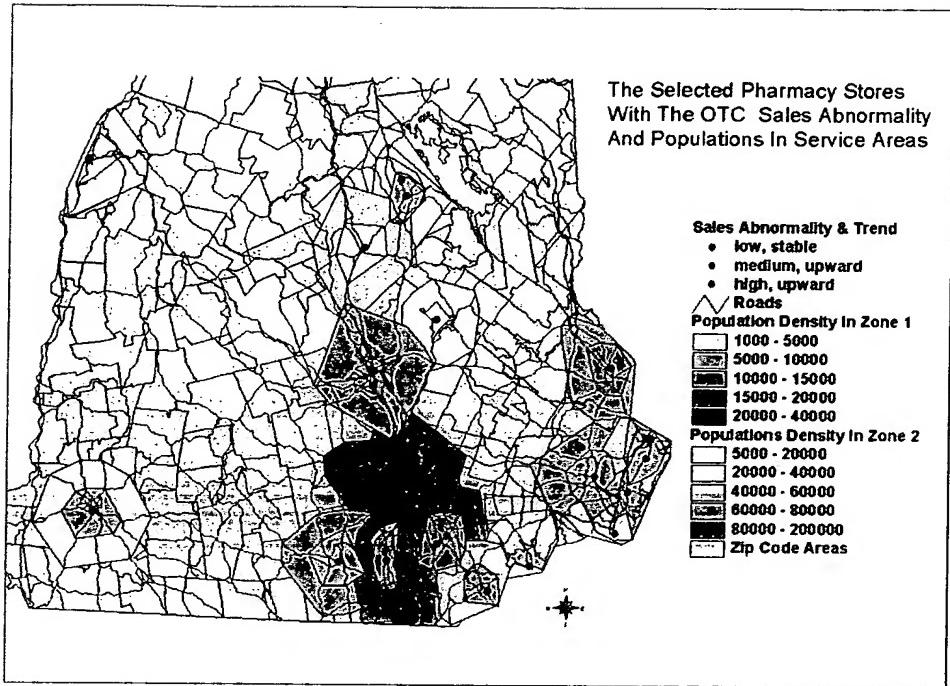
**Figure 9.** **An example of the categorized OTC sales abnormality analysis with population density in service area as the potential impact**

**Table 1. The validated state transitions from a State Si(k) to State Sj(k+1)**

| | $S_h(k+1)$ | $S_c(k+1)$ | $S_s(k+1)$ | $S_u(k+1)$ | $S_p(k+1)$ | $S_d(k+1)$ | $S_e(k+1)$ |
|---|---|---|---|---|---|---|---|
| $S_h(k)$ | $R_{h,h}$ | $R_{h,c}$ | 0 | 0 | 0 | 0 | 0 |
| $S_c(k)$ | $R_{c,h}$ | $R_{c,c}$ | $R_{c,s}$ | 0 | 0 | 0 | 0 |
| $S_s(k)$ | 0 | $R_{s,c}$ | 0 | $R_{s,u}$ | 0 | 0 | 0 |
| $S_u(k)$ | 0 | 0 | $R_{u,s}$ | $R_{u,u}$ | $R_{u,p}$ | 0 | 0 |
| $S_p(k)$ | 0 | 0 | 0 | 0 | 0 | $R_{p,d}$ | 0 |
| $S_d(k)$ | 0 | 0 | 0 | $R_{d,u}$ | 0 | $R_{d,d}$ | $R_{d,e}$ |
| $S_e(k)$ | $R_{e,h}$ | 0 | 0 | 0 | 0 | 0 | 0 |

Equation (1), (2) and (3):

If,
- *l* denotes the year
- L is the current year
- *m* is the number of years with historical data available ⸳
- *i* denotes the day of the month
- *j* is the month of a year. (an exemplary time unit used for simplified demonstration purposes only)

Thus, *i∈[1,31], j∈[1,12], l and m are integers.*

$$\bar{x}_j = \frac{1}{m x n} \sum_{l=L-1}^{L-m} \sum_{i=1}^{n} x_{l,j,i} \qquad (where \ i \in[1,31] \ and \ j \in[1,12]) \qquad (1)$$

$$s_j = \frac{1}{(m x n - 1)} \sqrt{\sum_{l=L-1}^{L-m} \sum_{i=1}^{n} (x_{l,j,i} - \bar{x}_j)^2} \qquad (where \ i \in[1,31] \ and \ j \in[1,12]) \qquad (2)$$

$$C_j = \bar{x}_j + t_{mxn-1} \ s_j \qquad (where \ i \in[1,31] \ and \ j \in[1,12]) \qquad (3)$$

Equation (4), (5) and (6):

$$d_{L,j,i} = \frac{x_{L,j,i} - \overline{x}_j}{\overline{x}_j} \times 100 \qquad (L - year, j - month, i - day\,of\,month) \qquad (4)$$

$$w_{L,j,i} = \sum_{n=0}^{n=6} d_{L,j,i-n} \qquad (7 - days\ cumulated\ deviation) \qquad (5)$$

$$v_{L,j,i} = d_{L,j,i} - d_{L,j,i-1} \qquad (change\,of\,the\,daily\ deviation) \qquad (6)$$

Equation (7), (8) and (9):

$$\text{supp}(\alpha(k)) = \{d_{l<L,j,i} \, ; \, F(d) \geq (1-\alpha)\} \tag{7}$$

$$\text{supp}(\beta(k)) = \{w_{l<L,j,i} \, ; \, F(w) \geq (1-\beta)\} \tag{8}$$

$$\text{supp}(\delta(k)) = \{v_{l<L,j,i} \, ; \, F(v) \geq (1-\delta)\} \tag{9}$$

Equation (10), (11), (12), (13), (14) and (15):

$$S_j(k+1) \Leftarrow \{S_i(k) \otimes R_{i,j} \otimes X_i(k), \ S_i(k) = k_w \, w_i(k)\} \qquad (10)$$

$$X_i(k) \Leftarrow B_{i,m}(\alpha(k), \beta(k), \delta(k)) \otimes \begin{bmatrix} d_{i,m}(k) \\ w_{i,m}(k) \\ v_{i,m}(k) \end{bmatrix} \qquad (11)$$

$$Y_i(k) \Leftarrow H_{i,n} \{(\gamma_0(k), \gamma_1(k), ..., \gamma_n(k)) \otimes \begin{bmatrix} S(k) \\ S(k-1) \\ ... \\ S(k-n) \end{bmatrix} \otimes G_i(k)\} \qquad (12)$$

*where*

$$X \subset \bigcup_{z=1}^{z} \text{supp}(z) \qquad (13)$$

$$\text{supp}(z) = \text{supp}(X_{z,1}) \times \text{supp}(X_{z,2}) \times \text{supp}(X_{z,3}) \qquad (14)$$

$$Y_i = (Q_{i,h}, T_{i,h}, P_{i,h}) \qquad (15)$$